

EXHIBIT 151

Case No. 3:23-cv-03417-VC
Attorney's Eyes Only

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA

RICHARD KADREY, an individual, et
al.

v.

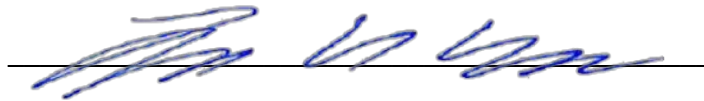
META PLATFORMS, INC., a Delaware
corporation;

Defendant.

Case No. 3:23-cv-03417-VC

EXPERT REPORT OF LYLE UNGAR, PH.D.

Signed in Philadelphia, Pennsylvania on January 10, 2025



Lyle Ungar, PhD

significant portions of texts that appear only a few times in their training data. In the case of Llama models, Meta's internal experiments (discussed below) and my own experimentation show that Llama does not "memorize" books authored by plaintiffs to this case and is highly unlikely to "memorize" similar works.

198. The term "memorization" is a misnomer, as LLMs, and Llama specifically, do not function by memorizing training data. Rather, Llama learns statistical patterns and relationships from the training data, generating predictions by assigning probabilities to possible next tokens based on the provided context, and using these probabilities to predict the next token. Unlike a fixed memory lookup, LLMs rely on a sampling strategy to produce next token predictions, meaning there is no single deterministic sequence of tokens attributed to a model. As discussed in **Section V**, for the continuation experiments conducted, greedy sampling, which selects the highest likelihood next token, is used to produce consistent and replicable results. However, as described in **Section IV.B.1.c**, commercial LLMs typically do not use greedy sampling, and will instead use a non-deterministic sampling technique such as top-p sampling which employ variability in the next token selection. For such sampling techniques, a change in just one token can completely alter the trajectory of subsequent predictions by reshaping the output probabilities.

199. "Memorization" occurs, as in the examples I describe below, when LLMs overfit on the average probabilities of specific training data. LLMs tend to memorize facts, phrases, and texts that appear very frequently in their training data and are thus frequently useful for next token prediction. For example, Llama is trained on many texts that contain quotations from the U.S. Constitution. Because text from the Constitution is found thousands of times in its training data, Llama's weights are slowly adjusted during training such that it usually correctly predicts the next token in sequences from the Constitution. The same is true of other very well-known and frequently repeated texts, such as the Gettysburg Address, the beginnings of famous books (*Moby Dick*, *Frankenstein*, and others), well-known poems (*The Road Not*

Taken, for example), and other frequently repeated texts that appear in multiple places on the Internet and thus are repeated multiple times in the training datasets.²⁸⁹

200. Researchers have extensively studied LLMs' memorization abilities and behavior to quantitatively and qualitatively describe the content LLMs memorize. Many measure ***discoverable memorization***, which checks whether the LLM completes a text verbatim when prompted with the beginning of the text.^{290, 291} Under discoverable memorization, texts are considered memorized if, when provided the first 50-100 tokens in a text, the LLM accurately reproduces the next 50 tokens from that text.²⁹² Discoverable memorization is favored because it is an efficient method of testing whether a specific text is memorized, which allows researchers to study memorization behavior by type and frequency of text. To test an LLM's rate of discoverable memorization researchers sample passages from the LLM's training data and prompt the LLM with the beginnings of those passages, much same as the continuations experiments described in **Section V**. As above, the LLM's outputs are compared with the true continuation of the text; if the model accurately continues the text for a predetermined number of tokens (often 50), the text is considered memorized.

201. Results from experiments measuring discoverable memorization show that LLMs memorize frequently repeated texts drastically more often than rarer texts. For example, when evaluating the PaLM LLM, engineers at Google tested its memorization rate on texts that

²⁸⁹ Other examples include texts frequently repeated on the internet for one reason or another. For example, website licenses and boilerplate terms and conditions statements, which can be repeated thousands of times in a dataset of web pages.

²⁹⁰ Jamie Hayes et al., "Measuring Memorization through Probabilistic Discoverable Extraction" (arXiv, October 25, 2024), <https://doi.org/10.48550/arXiv.2410.19482>.

²⁹¹ Other definitions of LLM memorization exist, including extractable memorization, which considers a text memorized if it can be produced by the LLM in response to any prompt, not just the beginning of the original text. To concretize the distinction, suppose that in response to the prompt "Write a speech memorializing the battle of Gettysburg," an LLM output the text of the Gettysburg Address. The Gettysburg Address would be considered extractably memorized. On the other hand, if in response to the prompt "Four score and seven years ago," an LLM completed the text of the Gettysburg Address, the Gettysburg Address would be considered discoverably memorized. Texts can be both extractably and discoverably memorized. Extractable memorization is much more difficult to test, however, because it is difficult to determine what prompt might elicit a specific text as output. To determine whether a specific text is memorized, testing for discoverable memorization is the simplest and most direct method. See: Jamie Hayes et al., "Measuring Memorization through Probabilistic Discoverable Extraction" (arXiv, October 25, 2024), <https://doi.org/10.48550/arXiv.2410.19482>.

²⁹² Jamie Hayes et al., "Measuring Memorization through Probabilistic Discoverable Extraction" (arXiv, October 25, 2024), <https://doi.org/10.48550/arXiv.2410.19482>.